Q1) Given standard deviation for 'n' distinct values $\{x_i\}_{i=1}^{n}$ is σ.

Mean $= \mu$

We know standard deviation $\sigma^2 = \dfrac{\sum_{i=1}^{n}(x_i-\mu)^2}{n-1}$

$\therefore \sum_{i=1}^{n}(x_i-\mu)^2 = \sigma^2(n-1) \longrightarrow (1)$

$\forall i \in [1,n]$ $(x_i-\mu)^2 \leq \sum_{i=1}^{n}(x_i-\mu)^2$ (as L.H.S is a part of Sum in R.H.S

By applying square root on both sides

$\forall i \in [1,n]$ $|x_i-\mu| \leq \sqrt{\sum_{i=1}^{n}(x_i-\mu)^2}$

$\therefore \forall i \in [1,n]$ $|x_i-\mu| \leq \sqrt{\sigma^2(n-1)}$ (from eq (1))

$\therefore \forall i \in [1,n]$ $|x_i-\mu| \leq \sigma\sqrt{n-1}$

2) Given 'μ' is the mean and 'τ' is the median.

Let take L.H.S $|\mu - \tau|$

$$\mu = \frac{x_1 + x_2 + \dots + x_n}{n}$$

$$|\mu - \tau| = \left| \frac{x_1 + x_2 + \dots + x_n}{n} - \tau \right|$$

$$= \left| \frac{x_1 + x_2 + \dots + x_n - n\tau}{n} \right|$$

$$= \left| \frac{(x_1 - \tau) + (x_2 - \tau) + \dots + (x_n - \tau)}{n} \right|$$

$$|\mu - \tau| = \frac{1}{n} \sum_{i=1}^{n} |x_i - \tau|$$

We know that $\sum_{i=1}^{n} |x_i - y|$ is minimum for $y = \tau$

$\therefore \sum_{i=1}^{n} |x_i - \tau| \leq \sum_{i=1}^{n} |x_i - \mu|$ (for any 'y', Let take $y = \mu$)

$$\therefore |\mu - \tau| = \frac{1}{n} \sum_{i=1}^{n} |x_i - \tau| \leq \frac{1}{n} \sum_{i=1}^{n} |x_i - \mu|$$

$$\therefore |\mu - \tau| \leq \frac{1}{n} \sum_{i=1}^{n} |x_i - \mu| \longrightarrow (1)$$

Let take R.H.S 'σ'

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}$$

Assume $|x_i - \mu| = y_i \quad \forall\ i \in [1, n]$

$$\therefore \sigma^2 = \frac{\sum_{i=1}^{n} y_i^2}{n}$$

eq (1) $\Rightarrow$ $|\mu - \tau| \leq \dfrac{\sum_{i=1}^{n} y_i}{n}$

Let 'σ_1^2' be the Variance of y_i for $i \in \{1, \dots n\}$

$$\sigma_1^2 = \frac{\sum_{i=1}^{n}(y_i - \mu_1)^2}{n} \quad (\mu_1 \text{ be mean of } y_i\text{'s})$$

as $\sigma_1^2 \geq 0 \Rightarrow \sum_{i=1}^{n}(y_i - \mu_1)^2 \geq 0$

$$\Rightarrow \sum_{i=1}^{n} y_i^2 - 2\mu_1 \sum_{i=1}^{n} y_i + n(\mu_1^2) \geq 0$$

$$\Rightarrow \sum_{i=1}^{n} y_i^2 - 2\mu_1(n\mu_1) + n\mu_1^2 \geq 0 \quad \left(\text{as } \mu_1 = \frac{\sum_{i=1}^{n} y_i}{n}\right)$$

$$\Rightarrow \sum_{i=1}^{n} y_i^2 \geq n\mu_1^2$$

$$\Rightarrow \sum_{i=1}^{n} y_i^2 \geq n\left(\frac{\sum_{i=1}^{n} y_i}{n}\right)^2$$

$$\Rightarrow \frac{\sum_{i=1}^{n} y_i^2}{n} \geq \left(\frac{\sum_{i=1}^{n} y_i}{n}\right)^2 \longrightarrow ③$$

we know $|\mu - \tau|^2 \leq \underbrace{\left(\frac{\sum_{i=1}^{n} y_i}{n}\right)^2 \leq \frac{\sum_{i=1}^{n} y_i^2}{n}}_{\text{from } ③} = \sigma^2$

$$\therefore \ |\mu - \tau|^2 \leq \sigma^2$$

$$\therefore \ \underline{|\mu - \tau| \leq \sigma}$$

Hence proved

Q 3) $P(C_i) = 1/3$, $i \in \{1,2,3\}$ also

probability that contestant choses door 1, door 2, door 3 are equal.

$\therefore P(Z_i) = 1/3,\ i \in \{1,2,3\}$

(a) Two events that contestant choses door and car is behind doors are independent.

$\therefore P(C_i|Z_1) = P(C_i) \quad \forall\ i \in \{1,2,3\}$

$\therefore P(C_1/Z_1) = P(C_1) = 1/3$

$P(C_2|Z_2) = P(C_2) = 1/3$

$P(C_3|Z_3) = P(C_3) = 1/3$

(b) Case 1 :- $i=1$

$\Rightarrow P(H_3|C_1, Z_1)$ = probability that host opens door 3 if car behinds door 1 & contestant chooses door 1

→ As car is behind door 1 and contestant chooses door 1 host should open door 2 and 3 with same probability (As given in Question)

$\therefore P(H_2|C_1, Z_1) = P(H_3|C_1, Z_1)$

and $P(H_1|C_1 Z_1) = 0$

$\therefore P(H_3|C_1 Z_1) = 1/2$

Case 2 :- $i=2$

$\Rightarrow P(H_3|C_2, Z_1)$ = probability that host opens door 3 if car is behind door 2 & contestant chooses door 1

→ As given in Question host does not choose to open either door 2 (or) 1

$\therefore$ He only opens door 3

$$\therefore \quad P(H_3 | C_2, Z_1) = 1$$

Case 3 :- $i = 3$

$P(H_3 | C_3, Z_1)$ = probability that host opens door 3 if car is behind door 3 and contestant opens door 1.

→ As given in Question host does not open door 3 as contestant choose door 3

$$\therefore \quad P(H_3 | C_3, Z_1) = 0$$

(c) Given $P(C_2 | H_3, Z_1) = \dfrac{P(H_3 | C_2, Z_1) \cdot P(C_2, Z_1)}{P(H_3, Z_1)}$

$P(H_3 | C_2, Z_1) = 1$ (done in b)

$P(C_2, Z_1) = \frac{1}{3} \times \frac{1}{3}$ (~~done in a~~ independent events)

$= 1/9$ (as $P(C_2, Z_1) = P(C_2) \cdot P(Z_1)$)

we have to find $P(H_3, Z_1)$

$$P(H_3, z_1) = P(H_3, z_1, c_1) + P(H_3, z_1, c_2) + P(H_3, z_1, c_3)$$

$$P(H_3, z_1) = P(H_3 | z_1, c_2)\, P(z_1, c_1) + P(H_3 | z_1, c_2)\, P(z_1, c_2) + P(H_3 | z_1, c_3)\, P(z_1, c_3)$$

$$P(H_3, z_1) = \frac{1}{2} \times \frac{1}{3} \times \frac{1}{3} + 1 \times \frac{1}{3} \times \frac{1}{3} + 0 = \frac{1}{6}$$

$$\left(\text{as } P(z_i, c_i) = P(z_i)\, P(c_i)\right)$$

$$\therefore \quad P(c_2 | H_3, z_1) = \frac{1 \times \frac{1}{9}}{\frac{1}{6}} = \frac{6}{9} = \frac{2}{3}$$

$$\therefore \quad P(c_2 | H_3, z_1) = \frac{2}{3}$$

(d)

$$P(C_i \mid H_3, Z_1) = \frac{P(H_3 \mid C_1, Z_1) \cdot P(C_1, Z_1)}{P(H_3, Z_1)}$$

$P(H_3 \mid C_1, Z_1) = \frac{1}{2}$ (as done in 'b')

$P(C_1, Z_1) = P(C_1) \times P(Z_1)$ (independent)

$= \frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$

$P(H_3, Z_1) = \frac{1}{6}$ (as done in 'c')

$\therefore P(C_1 \mid H_3, Z_1) = \frac{\frac{1}{2} \times \frac{1}{9}}{1/6} = \frac{6}{18} = \frac{1}{3}$

$\therefore P(C_1 \mid H_3, Z_1) = \frac{1}{3}$

Q 4) for median, relative mean squared error = 0.9634

for mean, relative mean squared error = 0.6481

for f = 30% :- for Quartile, relative mean squared error = 1.2824

∴ Mean has least relative mean squared error

for f = 60% :

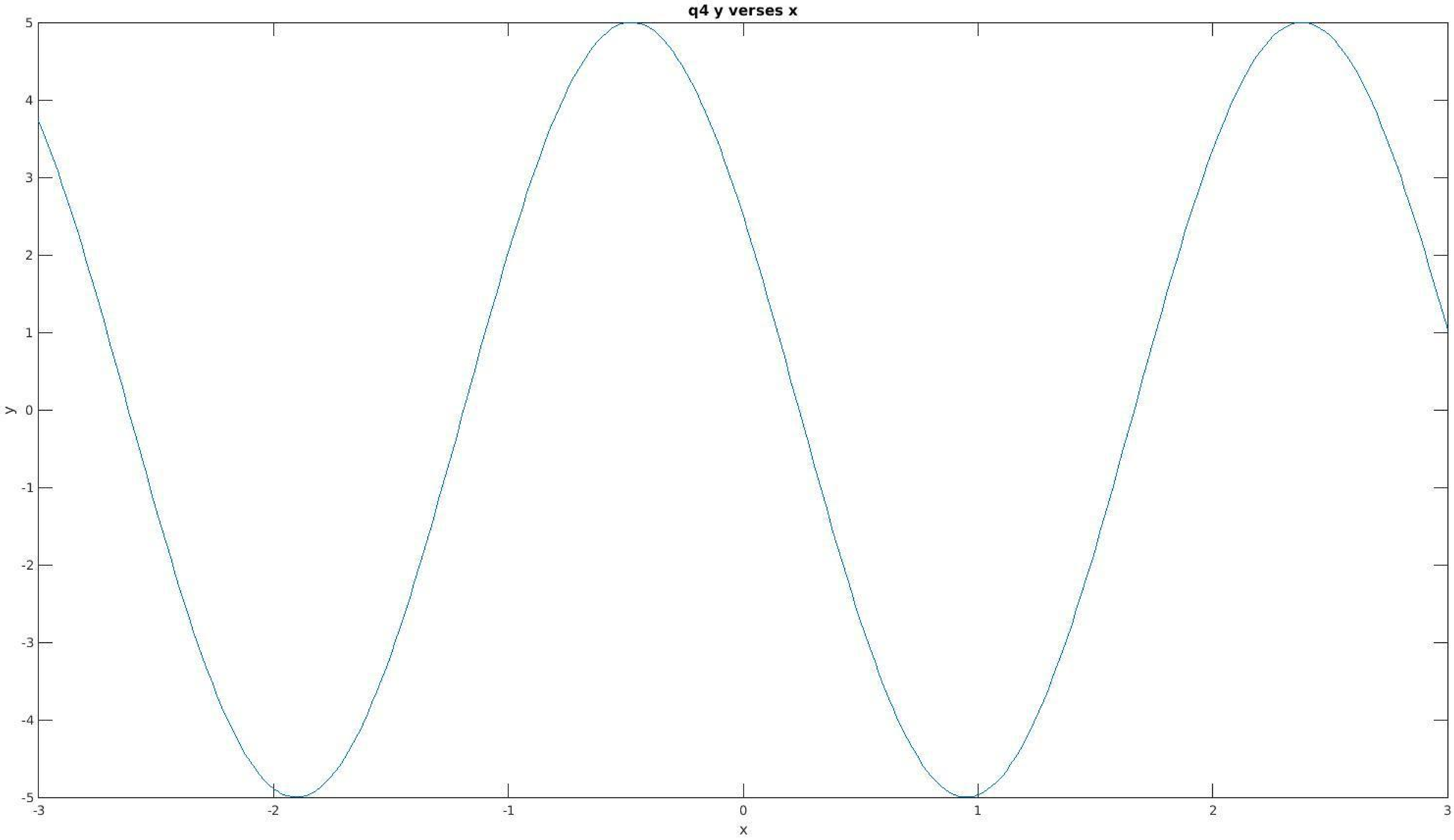
for median, Relative MS error = 0.6113

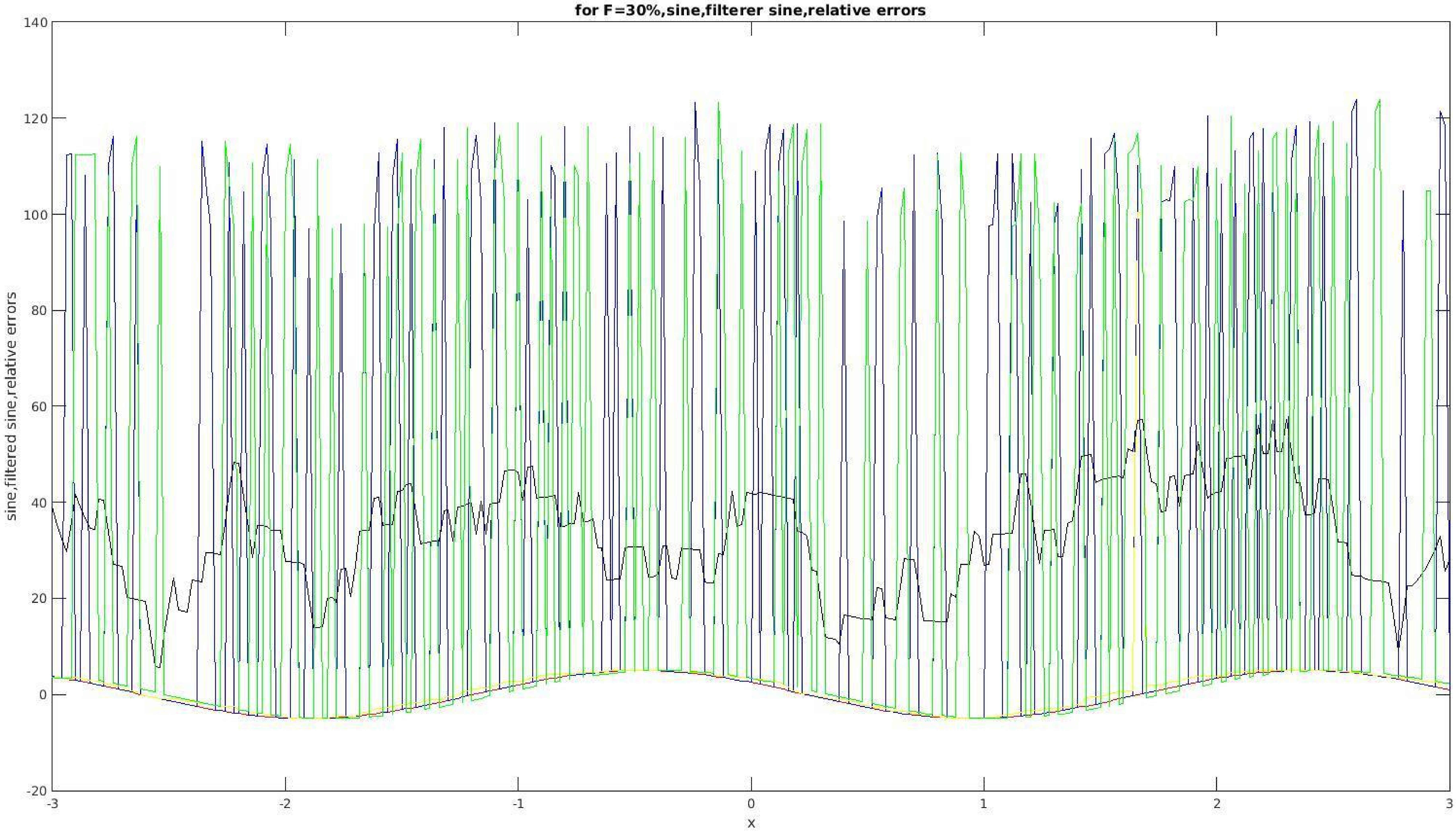
for mean, Relative MS error = 0.3815

for Quartile, Relative MS error = 0.7853

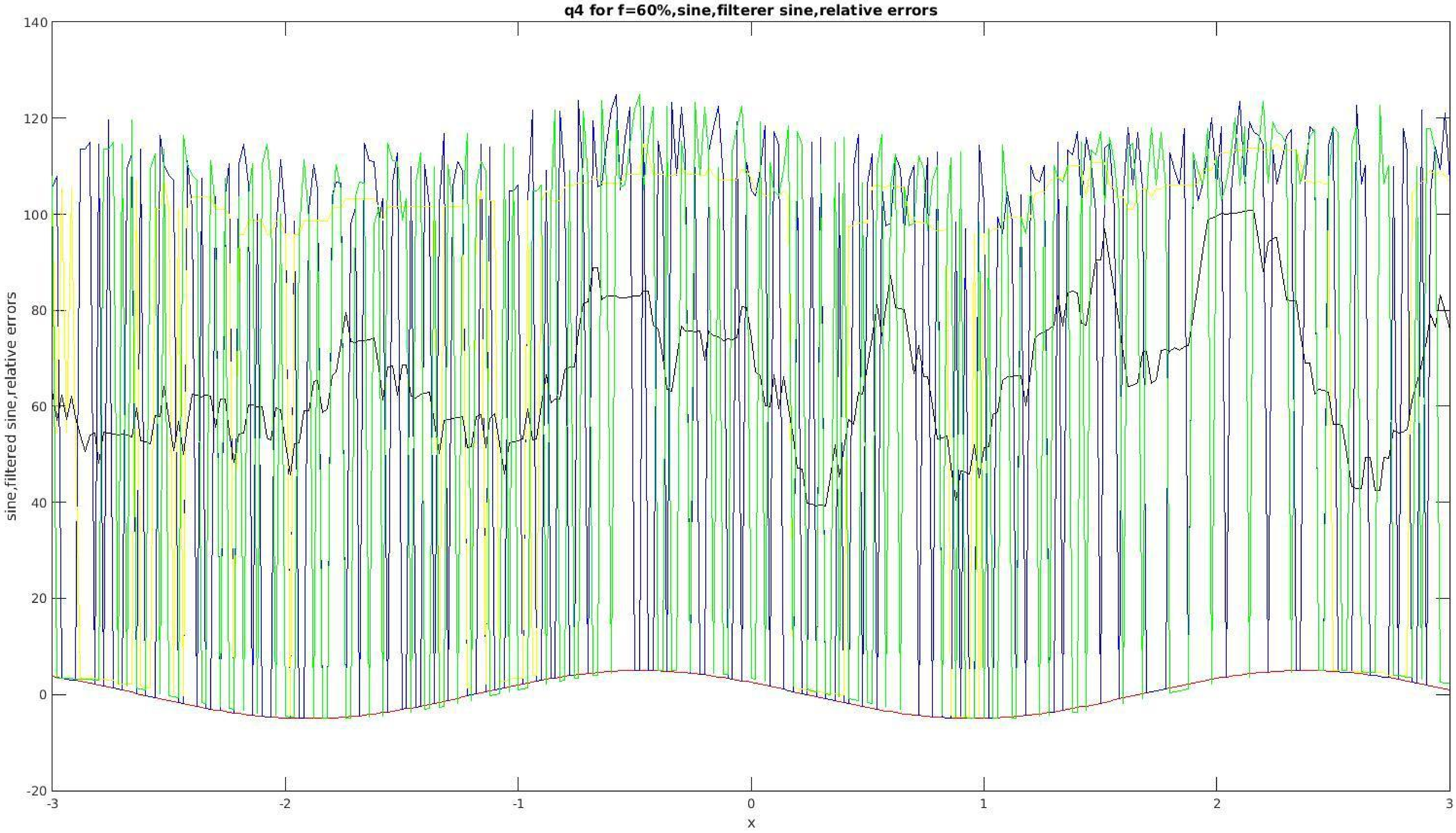
∴ Mean has least Relative mean squared error

* → Mean has low relative mean squared error because it takes into consideration of all values of data whereas Quartile and median ~~mostly~~ depend only on number of observations.


q4 y verses x
y
x
5
4
3
2
1
0
-1
-2
-3
-4
-5
-3
-2
-1
0
1
2
3



for F=30%,sine,filterer sine,relative errors
sine,filtered sine,relative errors
x
140
120
100
80
60
40
20
0
-20
-3
-2
-1
0
1
2
3


q4 for f=60%,sine,filterer sine,relative errors
sine,filtered sine,relative errors
x
140
120
100
80
60
40
20
0
-20
-3
-2
-1
0
1
2
3

Q5) a) old mean = $\frac{\sum_{i=1}^{n} x_i}{n}$

new sum of all values = old mean × n + new data value

New mean = $\frac{\text{old mean} \times n + \text{New data value}}{n+1}$

b) Update median:

Case a: (n is even) and (New data value < old Median)

Subcase A: New data value < $A(n/2)$

new median = $A(n/2)$

Subcase B: New data value ≥ $A(n/2)$

new median = New Data value

Case B: (n is odd) and (New data value < old median)

Sub case A: New data value < $A\left(\frac{n-1}{2}\right)$

new median = $\frac{A\left(\frac{n+1}{2}\right) + A\left(\frac{n+1}{2} - 1\right)}{2}$

Subcase B: else

new median = $\left[A\left(\frac{n+1}{2}\right) + \text{new data value}\right] / 2$

Case C: (n is even) and (New value > old median)

Sub case A: New value > $A(n/2 + 1)$

new median = $A(n/2 + 1)$

Subcase B: else

new median = New Data value

Case D: (n is odd) and (New value > old median)

Sub case A: New value > $A\left(\frac{n+1}{2}+1\right)$

$$\text{new median} = \frac{A\left(\frac{n+1}{2}\right) + A\left(\frac{n+1}{2}+1\right)}{2}$$

Sub case B:

else

$$\text{new median} = \left[\frac{A\left(\frac{n+1}{2}\right) + \text{New value}}{2}\right]$$

Case E: New Data value == old median

New median = old median

c) New standard deviation:

$$\sigma^2 = \frac{\Sigma(x - x_m)^2}{n-1}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n}[x_i]^2}{n-1} - \frac{2(\Sigma x_i)(x_m)}{(n-1)} \frac{\times n}{n} + \frac{n(x_m)^2}{n-1}$$

$$= \frac{\sum_{i=1}^{n}(x_i)^2}{n-1} - \frac{2(x_m)^2 \times n}{n-1} + \frac{n(x_m)^2}{n-1}$$

$$= \frac{\sum_{i=1}^{n} x_i^2 - n(x_m)^2}{n-1}$$

$$\sum_{i=1}^{n} x_i^2 = (n-1)\sigma^2 + n(x_m)^2$$

new sum of squares

$$\sum_{i=1}^{n} x_i = (n-1)\sigma^2 + (n)(x_m)^2 + (\text{new data value})^2$$

$$\sigma_2^2 = \frac{\sum_{i=1}^{n+1} x_i}{n} - \frac{(n+1)(x_{new\ mea})^2}{n}$$

$$\sigma = \sqrt{\left[\frac{(n-1)\sigma^2 + n(x_m)^2 + (new\ data)^2}{n}\right] - \frac{(n+1)(x_{new\ mean})^2}{n}}$$

Q6)

P(2 birthdays does not match)

↓

propability

= 1 - P(birthday's match)

P(2 birthday's match) = 1 - P(2 birthday's doesnot match)

↓

Let it be p

$$\Rightarrow p = 1 - \frac{(365)(365-1)\cdots(365-n+1)}{(365)^n} \longrightarrow ①$$

(as probability 2 birthday's does not match for 'n' persons) $= \dfrac{365 \cdot 364 \cdots (365-n+1)}{(365)^n}$

By apporxiamation :

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots$$

$$e^{-x} = 1 - \frac{x}{1!} + \frac{x^2}{2!} \cdots$$

$$\therefore 1 - \frac{1}{365} \simeq e^{-1/365}$$

Illy $1 - \dfrac{n}{365} \simeq e^{-n/365}$ for small 'n'

$\therefore$ from eq ①, $1 - p = \dfrac{(365)(364)\cdots(365-n+1)}{(365)^n}$

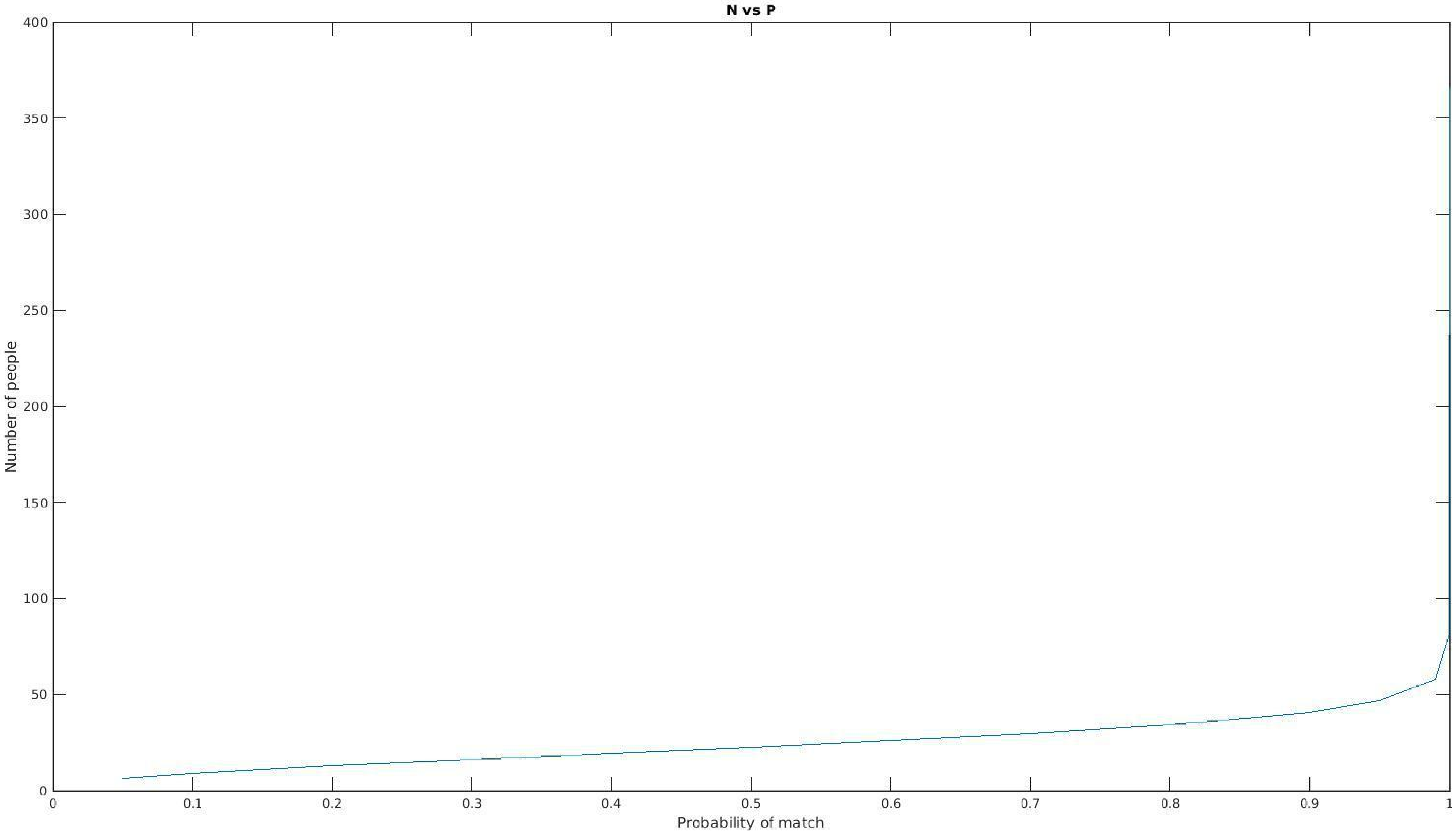
$$(1-p) = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right)\cdots\left(1 - \frac{n-1}{365}\right)$$

$$1 - p \simeq e^{-1/365} \times e^{-2/365} \times \cdots e^{-\frac{(n-1)}{365}}$$

$$\Rightarrow 1 - p \simeq e^{-\frac{(n-1)n}{365 \times 2}} \simeq e^{-\frac{n^2}{730}}$$

$$\therefore n^2 \simeq (-730)\log(1-p)$$

$$\therefore \boxed{n \simeq \sqrt{(-730)\log(1-p)}}$$


N vs P
Number of people
Probability of match
0
50
100
150
200
250
300
350
400
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1